

SAMPLE SOFTWARE WRITTEN

- **transformer-aiayn** - 2023: **Jax+Haiku** From scratch implementation of [Attention is All You Need](#) transformer, trained on WMT14/de-en using Cloud TPU, with novel attention entropy regularization. Achieved **Perplexity 4.95**, **Bleu 25.5** on newstest2013. [blog article](#), [talk](#)
- **motion-blur**: Custom CUDA Kernel for simulating photographic motion blur for arbitrary camera motion defined by homographic trajectories. **CUDA**, **C++** [blog article](#)
- **streamvis** - 2023: data logger + web client/server. Log metrics data during ML model training. auto-updating plots, post-logging configuration, multi-browser tabs, custom layouts. Uses single **protobuf** append-only log file, supports local or Cloud storage
- **simple-diffusion** - 2023: Implementation of Denoising Diffusion Probabilistic model from [original paper](#), using Radial Basis Function network, [blog article](#)
- **einsum-tuple** - 2022: Defines a proposed language for complete mathematical specification of tensor operations. Based on einsum, enhanced with index expressions and rank-agnostic indices, and an interpreter for validation
- **DRAW** - 2021: Repaired a popular broken **TensorFlow** implementation of DeepMind's LSTM-based autoregressive VAE. Provide analysis of training and mode collapse.
- **ae-wavenet** - 2020: Implemented VQ-VAE WaveNet to study unsupervised phoneme recognition and training dynamics. **173 stars**, **20 forks on github**, PyTorch-xla, TPU
- **depngs** - 2016: Designed novel statistically robust mutation detection in non-clonal DNA samples from next-gen sequence data. **5x faster** throughput than baselines. **C++**, **POSIX thread** queues.
- **wij-wemnan** - 2015: Designed data visualization [web app](#) with responsive user interaction and selection for 500k+ data points. **WebGL**, **Tensorflow.js**

EXPERIENCE

- **Glass Imaging, Inc** Los Altos, CA
Contractor *April 2024 - Sept 2024*
 - **Image Capture Pipeline**: Accelerated composite image generation synthetic data pipeline 8x, porting CPU to GPU code with **Vulkan**. Built file upload server in **go**, **Kotlin coroutines** based concurrency on phone app, and Python **asyncio** app with **Svelte** frontend.
- **Machine Learning Blog and Reading** San Francisco, CA
Independent Study *June 2017 - April 2024*
 - **Kernel Methods Tutorial** - 2021: Clarifies the relationship between input and function space through interactive **Svelte** and **SVG** based visualizations and mathematical derivations.
 - **VAE analysis** - 2021: An in-depth analysis of original VAE model explores implications of unimodal functions on training dynamics and generalization.
 - **To learn from actions, the brain must close the loop** - 2021: A mathematical justification for predictive coding as a necessity for credit assignment of agent action.
 - **Thoughts on Autoencoders** - 2022: I describe two surprising aspects of VAEs: a mathematical symmetry and a disclaimer about what the ELBO does not do, and explore implications.
 - **ML Reading and Courses** - 2017 - present: Deep Learning (Goodfellow), Linear Algebra (Shilov), Elements of Information Theory (Cover & Thomas), Probability Theory (Jaynes), Andrew Ng's course.
- **Config, Inc** San Francisco, CA
First employee, Chief Software Architect reporting to CEO *June 2019 - Oct 2019*
 - **Product Design**: Designed collaboration platform for Hardware design projects, with CAD versioning, parts tracking, approval flow. Implemented **SQL/pgSQL** Data Mart, **Rust** web server, **GraphQL + Pub/Sub**. Product [launched](#) in 2023

- **Amgen** South San Francisco, CA
Genome Analysis Unit *June 2010 - Dec 2016*
 Scientist Level 5
 - **Leukemia Study:** Discovered mechanism of Leukemia resistance to *blinatumomab* (commercialized as **BLINCYTO®**) in NGS responder study, applying my [mutation detection tool](#) to 30 TB genome data. **Results presented to CEO**
 - **Target Discovery:** Provided over **20 Oncologists** interactive visualization of genome-wide Cancer gene expression to discover drug targets. Built 100M data point **Oracle** Data Mart, **Spotfire** visualizations.
 - **Infrastructure Testing:** Conducted latency, memory bandwidth and throughput tests on bare metal and VM clusters and storage arrays for research group managing 200 TB genome data.
- **Broad Institute** Cambridge, MA
Computational R&D *May 2009 - June 2010*
 Computational Biologist
 - **HIV Immune Evasion Detection:** Designed Bayesian algorithm in C++ for reliable rare variant detection in non-clonal ultradeep next-gen HIV sequence data, applied to patient timepoints.

PROGRAMMING LANGUAGES AND TOOLS

- C++ (13y), Python (Jax, TensorFlow, PyTorch) (8y), SQL & PL/SQL (6y), Rust, CUDA, Javascript, O'Caml, Docker, Kubernetes, Google Cloud

LEADERSHIP AND AWARDS

- Awarded Distinction for Ph.D. Dissertation (**top 10%**)
- **98th** Percentile in Math Section of GRE exam
- Presented workshop: *Whole Genome Sequencing for Mutant Identification*. 17th International *C. elegans* Meeting. UCLA. June 2009. **1600 researchers** in attendance

EDUCATION

- **Columbia University** New York, NY
Department of Biochemistry and Molecular Biophysics *May 2007 - May 2009*
 Postdoctoral Research Scientist
 - **CisOrtho Transcription Factor Binding Site Search:** Provided genome-wide [search tool](#) (C++, MySQL) for *C. elegans* and *Drosophila* conserved sites, used by over **100 research labs**
- **Columbia University** New York, NY
Ph.D. Biochemistry & Molecular Biophysics *April 2007*
 - **Thesis Title:** *Membrane protein structure prediction with Bayesian Networks*
- **University of Chicago** Chicago, IL
Ben May Department for Cancer Research
 Research Associate
 - **Src Knockout Study:** Designed a plasmid vector and transfected into mouse Embryonic Stem cells, created a transgenic mouse, in order to study effects of gene knockout on mammary tumor progression.
- **UCLA** Los Angeles, CA
Department of Biochemistry and Molecular Biology
 Research Associate
 - **Protein Structure studies:** Attempted to solve atomic structure of σ^{54} transcription factor with Nuclear Magnetic Resonance spectra and X-ray crystallography to elucidate bacterial transcription.
- **Oberlin College** Oberlin, OH
B.A. Biology

Bigelow H, Doitsidou M, Sarin S, Hobert O. MAQGene: software to facilitate *C. elegans* mutant genome sequence analysis. **111 citations** *Nature Methods* **6(8)**, 549 (2009).

Bigelow H, Rost B. PROFtmb: a web server for predicting bacterial transmembrane beta barrel proteins. **73 citations** *Nucleic Acids Res* **1;34** 186-188 (2006).

Bigelow HR, Petrey DS, Liu J, Przybylski D, Rost B. Predicting transmembrane beta-barrels in proteomes. **210 citations** *Nucleic Acids Res* **32(8)**, 2566-2577 (2004).

Bigelow HR, Wenick AS, Wong A, Hobert O. CisOrtho: a program pipeline for genome-wide identification of transcription factor target genes using phylogenetic footprinting. **44 citations** *BMC Bioinformatics* **12;5** 27 (2004).